

The Psychometric Maze – Making Sense of the Technical Characteristics of Currently Available Assessments for Dual Language Learners

Roundtable Meeting on Supporting Positive Language and Literacy Development in Young Language Minority Children: Research, Policy and Practice

April 16-17th, 2008



Michael L. López, Ph.D.
National Center for Latino Child & Family Research
and



Sandra Barrueco, Ph.D.
The Catholic University of America



ELL Assessment Considerations

- **Important to consider whether goal is to assess at one point in time, capture growth over time, or both – primary research question of interest?**
- **Need to consider in which language(s) to assess, and how information will be used.**
- **Most measures administered in 1 language or the other, rarely in both, though dual administration or “conceptual scoring” likely may provide information on overall development.**
- **When selecting measures, careful attention must be paid to very specific psychometric properties of measures.**

Key questions to ask during selection and review process

- **Is there a match between the publishers' stated purpose and one's intended purpose?**
- **Are the domains or areas of interest captured through the items and scales?**
- **What is the depth and range of items and questions?**
- **How do and will young children respond to the measure?**
- **Who can/will administer it and for what purpose? How much training and time is needed?**

Key questions to ask during selection and review process

- **How well were young children included when developing the measure?**
- **What is the demographic composition of the normative samples across ethnicity, socioeconomic status, language, and region, among others?**
 - **How well does it match with your population?**
- **How strong are its psychometric properties (reliability and validity)?**
- **If a non-English measure, how strong are its cultural and linguistic measurement properties? How are its basic psychometric properties (reliability and validity)? How does these compare to the English version?**

Compendium of Measures for the Assessment of Young English Language Learners

(Barrueco, López, Ong, & Lozano, 2007)

- **Critical examination of the psychometric, linguistic, and cultural properties of the measures currently available for use with preschool-aged language minority children.**
- **Intent: Provide independent evaluations of measures that clinicians, teachers, or researchers are considering utilizing. The ultimate selection by these individuals is guided by the purpose for the assessment, along with the characteristics they need in a measure.**
- **Focused on direct language and literacy measures for use with Spanish-speaking preschoolers (did not include parent report measures or teacher measures for instructional purposes)**

Selection Process and Inclusion Criteria

(Barrueco, López, Ong, & Lozano, 2007)

■ **Step 1:**

- **Initial identification of > 1000 measures relating to early childhood, language, or literacy**
- **Multiple sources: MMY, publishing companies, articles, reports, and internet searches.**

■ **Step 2:**

- **Identified measures with:**
 - **target age range including 3-5 year olds,**
 - **ample coverage of language and/or literacy domains,**
 - **publication of English and Spanish forms, and**
 - **direct child assessment.**

18 Assessments Reviewed

- Battelle Developmental Inventory, Second Edition (BDI-2)
- Bilingual Vocabulary Assessment Measure
- Boehm-3 Preschool
- Brigance Screens
- Compton: Preschoolers Screening Evaluation
- Developmental Indicators for the Assessment of Learning-Third Edition (DIAL-3)
- Dynamic Indicators of Basic Early Literacy Skills 6th Edition (DIBELS)
- Early Literacy Skills Assessment (ELSA)
- Early Screening Inventory-Revised
- Expressive One-Word Picture Vocabulary Test (EOWPVT)
- FirstStep: Screening Test for Evaluating Preschoolers
- Merrill-Palmer-Revised Scales of Development (Merrill-Palmer-R)
- Peabody Picture Vocabulary Test-4 /Test de Vocabulario Imágenes Peabody (PPVT/TVIP)
- Pre-LAS 2000
- Preschool Language Scale-4 (PLS-4)
- Receptive One-Word Picture Vocabulary Test (ROWPVT)
- Woodcock-Muñoz Language Survey-Revised
- Young children's Achievement Test (YCAT)

Psychometric Equivalences

Reliability and Validity

- Reliability:
 - How precise or trustworthy a test score is in capturing the skills, attitudes or abilities it is purported to measure while
 - e.g., Cronbach's; split-half; Test-retest; inter-rater
- Validity
 - The degree to which all accumulated evidence supports the interpretation and use of test scores for a particular purpose
 - Face and Content
 - Internal Construct (Factor Analyses, etc.)
 - External Construct
 - Criterion (Concurrent, Predictive, Postdictive)

Cultural and Linguistic Considerations

(adapted from Bravo, 2003)

1. **Content equivalence**
 - **Are constructs & operational definitions pertinent for cultural group?**
 - Literature review, Expert panel, Observations, Interviews, Focus Groups
2. **Semantic and Cultural Equivalences of Item Translations**
 - **How was the measure translated and subsequently adjusted?**
 - **Upon translation, did the items conform to *semantic* or *content* equivalents?**
 - Translation/Back-Translation methods
 - Substitution: culturally concordant items or construct equivalents
 - Field Tests & Statistical analysis of item difficulty, order, etc.
 - Feedback from cultural informants: interviews, focus groups, panel comprised of community and experts

Cultural and Linguistic Considerations

(adapted from Bravo, 2003)

- **Structural Consistency across English and Spanish versions:**
 - **How similar are the items, length and format of the measure across the languages?**
- **Standardization:**
 - **Was the measure standardized with Spanish-speaking children?**
- **Technical Equivalence:**
 - **How similar is the reliability of the Spanish and English versions?**
- **Criterion Equivalence:**
 - **How similar is the validity of the Spanish and English versions?**
- **Conceptual Equivalence:**
 - **Is the same theoretical construct being assessed across the two languages?**

Example: Early Literacy Skills Assessment (ELSA)

- **Purpose:** To provide preschool programs and teachers with an authentic and meaningful way to assess young children's early literacy

- **Four early literacy domains:**
 - Comprehension
 - Phonological awareness
 - Alphabetic principle
 - Concepts about print

- **2 “Books”:** Dante Grows Up and Violet's Adventure

ELSA-Spanish

- **Content equivalence**
 - **Manual does not specifically discuss if and how relevance of four early literacy domains among Spanish-speaking children. However, there is research is support this.**
- **Semantic and Content Equivalence of Translations**
 - **The ELSA was translated and adapted to Spanish by one individual with background in both early childhood development and ESL instruction.**
- **Cultural Equivalency**
 - **The manual does not describe how any cultural discrepancies arose or were addressed.**
- **Structural Consistency across English and Spanish versions**
 - **The structure of the English and Spanish versions is consistent.**

Early Literacy Skills Assessment (ELSA): English

- > 1000 children in sample
- **Reliability**
 - **Test-retest reliability:** Not Described
 - **Internal consistency:** .80-96
 - **Inter-rater agreement:** Not Described
 - **Others (e.g., equivalent forms):** 2 forms statistically equivalent.
- **Validity**
 - **Face :** Not described, appears to measure area
 - **Content:** Literature-based; Procedure not described.
 - **Internal Construct:** Factor structures statistically supported
 - **External Construct :**
 - Younger children and those with disabilities score lower on ELSA *Violet's Adventure*, though not the case for Comprehension
 - Better for children over 45 months, particularly if one-time-point
 - **Criterion**
 - Moderate to strong correlations of related scales from the Woodcock-Johnson, Pre-CTOPP, and Get Ready to Read!.
 - Did not examine comprehension scale.

Early Literacy Skills Assessment (ELSA): Spanish

- > 300 children in sample
- **Reliability**
 - **Test-retest reliability:** Not Described
 - **Internal consistency:** .63-93
 - **Inter-rater agreement:** Not Described
 - **Others (e.g., equivalent forms):** Not examined.
- **Validity**
 - **Face :** Not described, appears to measure area
 - **Content:** Literature-based for English speakers; Procedure not described.
 - **Internal Construct:** Factor structures statistically supported; the phonological awareness subscale may be weaker than the others
 - **External Construct :**
 - Younger children and those with disabilities –unknown
 - Better for children over 45 months, particularly if one-time-point
 - **Criterion**
 - Not Examined

ELSA

□ **Technical Equivalence:**

- Overall, reliability is similar.
- The internal reliability of the phonological awareness and concepts about print subtests are lower for the Spanish version.
- The two forms of the ELSA (El Cambio en Dante and La Aventura de Violeta) were found to equivalent in English, but could not be examined for the Spanish version due to sample size.

□ **Criterion Equivalence:**

- Overall, validity is similar.
- Content validity was not described as thoroughly for the Spanish version, but likely adequate.
- Also, criterion validity has not yet been examined for the Spanish version.
- Both versions evidence similar factor structures, though the phonological awareness subscale may be weaker on the Spanish version.
- Both evidence floor effects for very young children.

□ **Conceptual Equivalence:**

- Overall, the Spanish and English versions appear to measure the same theoretical construct.

Overall Strengths and Weaknesses

□ **Strengths**

- One-on-one familiar reading session
- Internal construct validity more strongly established than for other measures, and appears similar to the English version.
- Alternate story forms

□ **Weaknesses**

- May work better for children over 45 months, particularly if engaging in a one-time assessment.
- Criterion validity of the ELSA-Spanish needs to be examined, along with the phonological awareness subscale.
- Raw scores can be transformed to a 3-skill level designation (Level 1: Early Emergent, Level 2: Emergent, Level 3: Competent Emergent), but validity of these levels is unknown.

Discussion

- **Consider assessment goals and research question(s) at hand when selecting measures.**
- **Adequate Measurement Selection/Development Process:**
 - **Carefully consider the reliability, validity, and linguistic and cultural equivalence across both English and Spanish versions.**
 - **Build in adequate measurement feasibility work to help guide final selection/refinement of measures.**
 - **Conduct adequate psychometric analyses with data that is collected.**
- **Keep hope alive!!!**



References

- **Barrueco, S., Lopez, M.L., Ong, C., & Lozano, P. (2007). *A Compendium of Measures for the Assessment of Young English Language Learners*. Report commissioned by First5 LA and Pew Charitable Trusts.**
- **Espinosa, L. & López, M.L. (2007). *Assessment Considerations for Young English Language Learners Across Different Levels of Accountability*. Commissioned paper for First 5 LA and the Pew Charitable Trusts' Early Childhood Accountability Project.**
- **Bravo, M. (2003) *Instrument Development; Cultural Adaptations for Ethnic Minority Research*. In Handbook of Racial & Ethnic Minority Psychology**

Acknowledgements

■ Funders:

- First 5 Los Angeles
- Pew Charitable Trusts' Early Childhood Accountability Project



■ Collaborators and Reviewers:

- Linda Espinosa, Ed.D., University of Missouri-Columbia
- First 5 LA: Patricia Lozano, M.A. & Christine Ong, Ph.D.
- CUA: Joanna Melon, Anusha Natajajaran, Saleem Hue Penny, & Stefanie Lockwood



The Psychometric Maze – Making Sense of the Technical Characteristics of Currently Available Assessments for Dual Language Learners

Roundtable Meeting on Supporting Positive Language and Literacy Development in Young Language Minority Children: Research, Policy and Practice



Michael L. López, Ph.D.
National Center for Latino Child & Family Research
(milopez@earthlink.net)

and



Sandra Barrueco, Ph.D.
The Catholic University of America